

On the typical values of the cross-correlation measure

László Mériai

Johann Radon Institute for Computational and Applied Mathematics

Austrian Academy of Sciences

Altenbergerstr. 69, 4040 Linz, Austria

e-mail: merai@cs.elte.hu

March 4, 2016

Abstract

Gyarmati, Mauduit and Sárközy introduced the *cross-correlation measure* $\Phi_k(\mathcal{F})$ to measure the randomness of families of binary sequences $\mathcal{F} \subset \{-1, 1\}^N$.

In this paper we study the order of magnitude of the cross-correlation measure $\Phi_k(\mathcal{F})$ for typical families. We prove that, for most families $\mathcal{F} \subset \{-1, 1\}^N$ of size $2 \leq |\mathcal{F}| < 2^{N/12}$, $\Phi_k(\mathcal{F})$ is of order $\sqrt{N \log \binom{N}{k} + k \log |\mathcal{F}|}$ for any given $2 \leq k \leq N/(6 \log_2 |\mathcal{F}|)$.

2000 Mathematics Subject Classification: Primary 11K45, 68R15

Key words and phases: pseudorandom, binary sequence, correlation measure, cross-correlation measure

1 Introduction

Recently, in a series of papers the pseudorandomness of *finite binary sequences* $E_N = (e_1, e_2, \dots, e_N) \in \{-1, 1\}^N$ has been studied. In particular measures of pseudorandomness have been defined and investigated; see [1, 3, 5, 7] and the references therein.

For example, Mauduit and Sárközy [7] introduced the *correlation measure of order k* $C_k(E_N)$ of the sequences E_N . Namely, for a k -tuple $D = (d_1, d_2, \dots, d_k)$ with non-negative integers $0 \leq d_1 < d_2 < \dots < d_k < N$ and $M \in \mathbb{N}$ with $M + d_k \leq N$ write

$$V_k(E_N, M, D) = \sum_{n=1}^M e_{n+d_1} e_{n+d_2} \dots e_{n+d_k}.$$

The final publication is available at Springer via <http://dx.doi.org/10.1007/s00605-016-0886-0>

Then $C_k(E_N)$ is defined as

$$C_k(E_N) = \max_{M,D} |V(E_N, M, D)| = \max_{M,D} \left| \sum_{n=1}^M e_{n+d_1} e_{n+d_2} \cdots e_{n+d_k} \right|.$$

Cassaigne, Mauduit and Sárközy [3] studied the typical values of $C_k(E_N)$, when the binary sequences E_N are chosen equiprobably from $\{-1, 1\}^N$. Later Alon, Kohayakawa, Mauduit, Moreira and Rödl [1] improved their result.

Theorem 1. *For fixed $0 < \varepsilon_0 \leq 1/16$, there is a constant $N_0 = N_0(\varepsilon_0)$ such that if $N \geq N_0$, then, with probability at least $1 - \varepsilon_0$, we have*

$$\begin{aligned} \frac{2}{5} \sqrt{N \log \binom{N}{k}} &< C_k(E_N) < \sqrt{(2 + \varepsilon_1) N \log \left(N \binom{N}{k} \right)} \\ &< \sqrt{(3 + \varepsilon_0) N \log \binom{N}{k}} < \frac{7}{4} \sqrt{N \log \binom{N}{k}}. \end{aligned}$$

for every integer k with $2 \leq k \leq N/4$, where $\varepsilon_1 = \varepsilon_1(N) = (\log \log N) / \log N$.

Recently, Schmidt [9] showed that for fixed k , the correlation measure C_k of order k converges strongly, and so has limiting distribution.

In order to study the pseudorandomness of families of finite binary sequences $\mathcal{F} \subset \{-1, 1\}^N$, Gyarmati, Mauduit and Sárközy [6] introduced the notion of the *cross-correlation measure*.

Let $N, k \in \mathbb{N}$, and for any k binary sequences $E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(k)}$ with

$$E_N^{(i)} = (e_1^{(i)}, e_2^{(i)}, \dots, e_N^{(i)}) \in \{-1, 1\}^N, \quad \text{for } i = 1, 2, \dots, k,$$

and any $M \in \mathbb{N}$ and k -tuple $D = (d_1, \dots, d_k)$ of non-negative integers with

$$0 \leq d_1 \leq d_2 \leq \cdots \leq d_k < M + d_k \leq N, \tag{1}$$

write

$$V_k(E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(k)}, M, D) = \sum_{n=1}^M e_{n+d_1}^{(1)} e_{n+d_2}^{(2)} \cdots e_{n+d_k}^{(k)}.$$

Let

$$\tilde{C}_k(E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(k)}) = \max_{M,D} \left| V_k(E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(k)}, M, D) \right|,$$

where the maximum is taken over all $D = (d_1, \dots, d_k)$ and $M \in \mathbb{N}$ satisfying (1) with the additional restriction that if $E_N^{(i)} = E_N^{(j)}$ for some $i \neq j$, then we must not have $d_i = d_j$. Then the *cross-correlation measure of order k* of the family \mathcal{F} of binary sequences $E_N \in \{-1, 1\}^N$ is defined as

$$\Phi_k(\mathcal{F}) = \max \tilde{C}_k(E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(k)}),$$

where the maximum is taken over all k -tuples of binary sequences

$$\left(E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(k)}\right), \quad E_N^{(i)} \in \mathcal{F}, \quad \text{for } i = 1, \dots, k.$$

Clearly, for family $\mathcal{F} = \{E_N\}$ of size 1 we have

$$\Phi_k(\{E_N\}) = C_k(E_N).$$

On the other hand for general \mathcal{F} we have

$$\Phi_k(\mathcal{F}) \geq \max_{E_N \in \mathcal{F}} C_k(E_N).$$

2 Typical values of $\Phi_k(\mathcal{F})$

In this paper we estimate $\Phi_k(\mathcal{F})$ for "random" families \mathcal{F} of sequences E_N with given length N and family size $|\mathcal{F}|$, i.e. we choose a family \mathcal{F} from all subsets of $\{-1, 1\}^N$ of size $|\mathcal{F}|$ with the same probability.

Clearly, the typical value of $\Phi_k(\mathcal{F})$ strongly depends on the size of the family \mathcal{F} . If \mathcal{F} is large: $|\mathcal{F}| > 2^{cN}$ with some $0 < c < 1/2$, then $\Phi_k(\mathcal{F}) \gg N$ ($c = 0.18$ can be chosen, see [6]). On the other hand, if $|\mathcal{F}| < 2^{cN}$ with $c \leq 1/12 = 0.0833\dots$, then the behavior of $\Phi_k(\mathcal{F})$ can be controlled.

Theorem 2. *For a given $\varepsilon > 0$, there exists N_0 , such that if $N > N_0$ and $1 \leq \log_2 |\mathcal{F}| < N/12$, then we have with probability at least $1 - \varepsilon$, that*

$$\frac{2}{5} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{F}| \right)} < \Phi_k(\mathcal{F}) < \frac{5}{2} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{F}| \right)}$$

for every integer k with $2 \leq k \leq N/(6 \log_2 |\mathcal{F}|)$.

The cross-correlation measure Φ can be also defined for *binary sequence generators* instead of families of sequences. Namely, let \mathcal{S} be a given set (set of parameters or seeds) and $N \in \mathbb{N}$ be an integer. A binary sequence generator is a map $G : \mathcal{S} \rightarrow \{-1, 1\}^N$ where

$$s \mapsto E_N(s) = (e_1(s), e_2(s), \dots, e_N(s)) \in \{-1, 1\}^N.$$

For a survey of (pseudorandom) sequence generators, in particular their application in cryptography, see [8, Chapters 5 and 6].

The cross-correlation measure of the generator G can be defined in the following way:

Let $M, k_1, k_2, \dots, k_\ell \geq 1$ be integers with the restriction $k = k_1 + k_2 + \dots + k_\ell \geq 2$. Let $D = (d_1^1, d_2^1, \dots, d_{k_1}^1, d_1^2, d_2^2, \dots, d_{k_2}^2, \dots, d_1^\ell, d_2^\ell, \dots, d_{k_\ell}^\ell)$ be a k -tuple such that

$$0 \leq d_1^i < d_2^i < \dots < d_{k_i}^i < M + d_{k_i}^i \leq N, \quad \text{for } i = 1, \dots, \ell. \quad (2)$$

Then for distinct $s_1, s_2, \dots, s_\ell \in \mathcal{S}$ write

$$\begin{aligned} & V_{k_1, k_2, \dots, k_\ell}(E_N(s_1), E_N(s_2), \dots, E_N(s_\ell), M, D) \\ &= \sum_{n=1}^M e_{n+d_1^1}(s_1) e_{n+d_2^1}(s_1) \dots e_{n+d_{k_1}^1}(s_1) \dots e_{n+d_1^\ell}(s_\ell) e_{n+d_2^\ell}(s_\ell) \dots e_{n+d_{k_\ell}^\ell}(s_\ell). \end{aligned}$$

The *cross-correlation measure of order k* of the generator G is defined as

$$\tilde{\Phi}_k(G) = \max |V_{k_1, k_2, \dots, k_\ell}(E_N(s_1), E_N(s_2), \dots, E_N(s_\ell), M, D)|,$$

where the maximum is taken over all integers $k_1, k_2, \dots, k_\ell \geq 1$ such that $k = k_1 + k_2 + \dots + k_\ell$, all $s_1, s_2, \dots, s_\ell \in \mathcal{S}$, and all M and D satisfying (2).

If the generator G is collision free (injection), then $\tilde{\Phi}_k(G) = \Phi_k(\mathcal{F})$ with the family

$$\mathcal{F} = \mathcal{F}(G) = \{E_N(s) : s \in \mathcal{S}\}.$$

On the other hand, if there is a collision: $E_N(s) = E_N(s')$ for $s \neq s'$, then $\tilde{\Phi}_k(G) = N$.

First, we estimate the value of $\tilde{\Phi}_k(G)$ for "random" generator G . For each $s \in \mathcal{S}$ and $1 \leq n \leq N$ we choose $e_n(s)$ from $\{-1, 1\}$ uniformly and independently. Then we have

Theorem 3. *For a given $\varepsilon > 0$, there exists N_0 , such that if $N > N_0$ and $1 \leq \log_2 |\mathcal{S}| < N/12$ then we have with probability at least $1 - \varepsilon$, that*

$$\frac{2}{5} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} < \tilde{\Phi}_k(G) < \frac{5}{2} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)}$$

for every integer k with $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$.

We can prove Theorem 2 as a corollary of Theorem 3.

Theorem 2. Throughout the proof we assume, that the integer N is large enough.

First we show, that for $|\mathcal{S}| < 2^{cN}$ with $0 < c < 1/2$, the probability of the collision is small:

$$\mathbb{P}(\nexists s, s' \in \mathcal{S} : E_N(s) = E_N(s')) = 1 - o(1) \quad (3)$$

Indeed, this probability is

$$\frac{\binom{2^N}{|\mathcal{S}|} \cdot |\mathcal{S}|!}{(2^N)^{|\mathcal{S}|}} = \left(1 - \frac{1}{2^N}\right) \cdot \left(1 - \frac{2}{2^N}\right) \dots \left(1 - \frac{|\mathcal{S}| - 1}{2^N}\right) \geq \left(1 - \frac{|\mathcal{S}|}{2^N}\right)^{|\mathcal{S}|}.$$

Since for all $0 < \delta < 1$ there is N_0 such that if $N \geq N_0$ we have

$$\left(1 - \frac{|\mathcal{S}|}{2^N}\right)^{|\mathcal{S}|} \geq \left(1 - \frac{1}{2^N/|\mathcal{S}|}\right)^{\delta 2^N/|\mathcal{S}|} \geq (e^{-\delta} + o(1)),$$

which gives (3).

Now let us assume, that Theorem 3 holds with ε_1 and let ε' be the probability of the collision. Then for a random generator G we have

$$\begin{aligned}
\varepsilon_1 &> \mathbb{P} \left(\tilde{\Phi}_k(G) > \frac{5}{2} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} \right) \\
&= \mathbb{P} \left(\tilde{\Phi}_k(G) > \frac{5}{2} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} \middle| \text{there is no collision} \right) \\
&\quad \cdot \mathbb{P}(\text{there is no collision}) \\
&\quad + \mathbb{P} \left(\tilde{\Phi}_k(G) > \frac{5}{2} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} \middle| \text{there is a collision} \right) \\
&\quad \cdot \mathbb{P}(\text{there is a collision}) \\
&= \mathbb{P} \left(\tilde{\Phi}_k(G) > \frac{5}{2} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} \middle| \text{there is no collision} \right) (1 - \varepsilon') \\
&\quad + 1 \cdot \varepsilon'.
\end{aligned}$$

If G is chosen uniformly from all generators with the condition that there is no collision, then the family $\mathcal{F} = \mathcal{F}(G)$ is uniformly distributed within all families of size $|\mathcal{F}| = |\mathcal{S}|$. Thus

$$\begin{aligned}
&\mathbb{P} \left(\tilde{\Phi}_k(G) > \frac{5}{2} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} \middle| \text{there is no collision} \right) \\
&= \mathbb{P} \left(\Phi_k(\mathcal{F}(G)) > \frac{5}{2} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{F}(G)| \right)} \right)
\end{aligned}$$

and so

$$\mathbb{P} \left(\Phi_k(\mathcal{F}) > \frac{5}{2} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{F}| \right)} \right) < \frac{\varepsilon_1 - \varepsilon'}{1 - \varepsilon'}.$$

In the same way we get

$$\mathbb{P} \left(\Phi_k(\mathcal{F}) < \frac{2}{5} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{F}| \right)} \right) > 1 - \frac{\varepsilon_1 - \varepsilon'}{1 - \varepsilon'}.$$

Choosing $\varepsilon = \frac{\varepsilon_1 - \varepsilon'}{1 - \varepsilon'}$ we get the result. \square

3 Estimates for $\tilde{\Phi}_k(G)$ for random generator G

In this section we consider G as a "random" generator i.e. $e_n(s)$ are independent and uniform random variables in $\{-1, 1\}$, for each $s \in \mathcal{S}$ and $1 \leq n \leq N$.

3.1 Estimates for the binomial distribution

The proof of Theorem 3 is based on estimations on tails of the binomial distribution. First we summarize some basic facts about their properties.

Let $S(n, p)$ be the sum of n independent Bernoulli random variables with mean p . First we state the following consequences of the de Moivre-Laplace theorem (see e.g. [4, Chapter 1, Theorem 6]) for $p = 1/2$.

Lemma 1. (i) For any $c = c(n) > 0$ with $c = o(n^{1/6})$, we have

$$\begin{aligned} \mathbb{P}\left(S(n, 1/2) \geq \left\lfloor \frac{n}{2} \right\rfloor + c\sqrt{n}\right) &= \sum_{\ell \geq c\sqrt{n}} \frac{1}{2^n} \binom{n}{\lfloor n/2 \rfloor + \ell} \\ &= \left(\sqrt{\frac{2}{\pi} + o(1)} \right) \left(\int_c^\infty e^{-2x^2} dx \right). \end{aligned} \quad (4)$$

In particular, if we further have that $c \rightarrow \infty$, then

$$\mathbb{P}\left(S(n, 1/2) \geq \left\lfloor \frac{n}{2} \right\rfloor + c\sqrt{n}\right) = \frac{e^{-2c^2}}{2c\sqrt{2\pi}}(1 + o(1)). \quad (5)$$

(ii) The estimates (4) and (5) hold for the lower tail

$$\mathbb{P}\left(S(n, 1/2) \leq \left\lfloor \frac{n}{2} \right\rfloor - c\sqrt{n}\right)$$

as well.

Let $\{x\} = x - \lfloor x \rfloor$. We have the following lower estimate for the symmetric binomial distribution (see [1, Fact 10]).

Lemma 2. Let n and c be integers with

$$-\left\lfloor \frac{n}{2} \right\rfloor \leq c \leq \left\lceil \frac{n}{2} \right\rceil.$$

If n is sufficiently large, then

$$\mathbb{P}\left(S(n, 1/2) = \left\lfloor \frac{n}{2} \right\rfloor + c\right) = \frac{1}{2^n} \binom{n}{\lfloor n/2 \rfloor + c} \geq (1 + o(1)) 2^{-4(c + \{n/2\})^2/n} \sqrt{\frac{2}{\pi n}}.$$

Let

$$S^\pm(n) = \sum_{1 \leq i \leq n} X_i,$$

where X_i ($1 \leq i \leq n$) are independent random variables with mean 0, that is,

$$\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = +1) = 1/2.$$

Clearly, $(S^\pm(n) + n)/2$ is binomially distributed with parameters n and $1/2$. The following lemma states a well-known estimate for large deviation of $S^\pm(n)$ (see e.g. [2, Appendix 2]).

Lemma 3. Let X_i ($1 \leq i \leq n$) be independent ± 1 random variables with mean 0. Let $S^\pm(n) = \sum_{1 \leq i \leq n} X_i$. For any real number $a > 0$, we have

$$\mathbb{P}(S^\pm(n) > a) < e^{-a^2/2n}.$$

3.2 Proof of Theorem 3

We prove Theorem 3 in two parts. First, we prove the upper estimate for $\tilde{\Phi}_k(G)$ for typical generator G .

Lemma 4. For $1 \leq \log_2 |\mathcal{S}| < \log_2 N$ we have

$$\tilde{\Phi}_k(G) < 2\sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)},$$

and for $\log_2 N \leq \log_2 |\mathcal{S}| < N/12$ we have

$$\tilde{\Phi}_k(G) < 2\sqrt{N \left(k \log N + \log \binom{|\mathcal{S}|}{k} \right)} < 2\sqrt{N \left(\log \binom{N}{k} + (1 + o(1))k \log |\mathcal{S}| \right)}$$

with probability tending to 1 as $N \rightarrow \infty$ for every integer k with $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$.

Proof. Assume first, that $1 \leq \log_2 |\mathcal{S}| < \log_2 N$.

Let us consider the event

$$V_{k_1, k_2, \dots, k_\ell}(E_N(s_1), E_N(s_2), \dots, E_N(s_\ell), M, D) > 2\sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} \quad (6)$$

for fixed integers $k, \ell, k_1, k_2, \dots, k_\ell, M$ and k -tuple D with restrictions $k = k_1 + k_2 + \dots + k_\ell$ and (2).

Since $e_n(s)$ are independent for $1 \leq n \leq N$ and $s \in \mathcal{S}$, then

$$z_n = e_{n+d_1^1}(s_1) e_{n+d_2^1}(s_1) \dots e_{n+d_{k_1}^1}(s_1) \dots e_{n+d_1^\ell}(s_\ell) e_{n+d_2^\ell}(s_\ell) \dots e_{n+d_{k_\ell}^\ell}(s_\ell)$$

are also independent and uniform in $\{-1, 1\}$. This follows from the observations, that for each j the sequence

$$\left(e_{1+d_1^j}(s_j) \dots e_{1+d_{k_j}^j}(s_j), \dots, e_{N-d_{k_j}^j+d_1^j}(s_j) \dots e_N(s_j) \right)$$

is uniformly distributed in $\{-1, 1\}^{N-d_{k_j}^j}$, and the sequence

$$(x_1(1), \dots, x_m(1), \dots, x_1(p), \dots, x_m(p))$$

is uniform in $\{-1, 1\}^{pm}$ if and only if

$$(x_1(1), \dots, x_m(1), x_1(p-1), \dots, x_m(p-1), x_1(1) \cdots x_1(p), \dots, x_1(p) \cdots x_m(p))$$

is uniform in $\{-1, 1\}^{pm}$.

Then

$$V_{k_1, k_2, \dots, k_\ell}(E_N(s_1), E_N(s_2), \dots, E_N(s_\ell), M, D)$$

has the same distribution as $S^\pm(M, 1/2)$. By Lemma 3 we have, that (6) holds with probability less than

$$\exp \left\{ -\frac{1}{2M} 4N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right) \right\} \leq \left(\binom{N}{k} |\mathcal{S}|^k \right)^{-2}.$$

Summing over all possible choices of $\ell, k_1, k_2, \dots, k_\ell, s_1, s_2, \dots, s_\ell, M$ and D we get

$$\begin{aligned} & \mathbb{P} \left(\tilde{\Phi}_k(G) > 2\sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} \right) \\ & \leq \sum_{\ell} \sum_{k_1, k_2, \dots, k_\ell} \sum_{s_1, s_2, \dots, s_\ell} \sum_M \sum_D \left(\binom{N}{k} |\mathcal{S}|^k \right)^{-2}. \end{aligned} \quad (7)$$

For $k \leq |\mathcal{S}|$ we estimate the number of k -tuples D by N^k . Thus (7) is less than

$$\begin{aligned} & \left(\binom{N}{k} |\mathcal{S}|^k \right)^{-2} \sum_{\ell=1}^k \binom{k-1}{\ell-1} \binom{|\mathcal{S}|}{\ell} N^{k+1} \\ & \leq 2^{k-1} \frac{\binom{|\mathcal{S}|}{k} N^{k+1}}{\left(\binom{N}{k} |\mathcal{S}|^k \right)^2} \leq 2^{k-1} \frac{\frac{e^k |\mathcal{S}|^k}{k^k} N^{k+1}}{\frac{N^{2k}}{k^{2k}} |\mathcal{S}|^{2k}} = \frac{2^{k-1} (ek)^k}{|\mathcal{S}|^k N^{k-1}} \leq e \left(\frac{2e}{N} \right)^{k-1} \\ & \leq \frac{2e^2}{N} \left(\frac{e}{3} \right)^{k-2}, \end{aligned} \quad (8)$$

where we used $\left(\frac{a}{b} \right)^b \leq \left(\frac{a}{b} \right) \leq \left(\frac{ea}{b} \right)^b$.

Next, consider (7) for $k > |\mathcal{S}|$. We estimate the number of k -tuples D of form (2) with the restriction $\ell \leq |\mathcal{S}|$ by

$$\binom{N}{k_1} \binom{N}{k_2} \cdots \binom{N}{k_\ell} \leq \frac{(eN)^{k_1}}{k_1^{k_1}} \frac{(eN)^{k_2}}{k_2^{k_2}} \cdots \frac{(eN)^{k_\ell}}{k_\ell^{k_\ell}} = \frac{(eN)^k}{e^{k_1 \log k_1 + k_2 \log k_2 + \cdots + k_\ell \log k_\ell}} \quad (9)$$

Since the function $x \log x$ (with $0 \log 0 = 0$) is convex, writing $k_{\ell+1} = \cdots = k_{|\mathcal{S}|} = 0$, we get by the Jensen inequality, that

$$k_1 \log k_1 + k_2 \log k_2 + \cdots + k_\ell \log k_\ell = \sum_{i=1}^{|\mathcal{S}|} k_i \log k_i \geq k \log \frac{k}{|\mathcal{S}|}.$$

Whence we get that (9) is less than

$$\frac{(eN)^k}{\left(\frac{k}{|\mathcal{S}|}\right)^k} \leq \binom{N}{k} (e|\mathcal{S}|)^k. \quad (10)$$

By (9) and (10) we have that (7) for $k > |\mathcal{S}|$ is less than

$$\begin{aligned} & \left(\binom{N}{k} |\mathcal{S}|^k \right)^{-2} \binom{N}{k} (e|\mathcal{S}|)^k \sum_{\ell} \sum_{k_1, k_2, \dots, k_{\ell}} \sum_{s_1, s_2, \dots, s_{\ell}} \sum_M 1 \\ & \leq \left(\binom{N}{k} |\mathcal{S}|^k \right)^{-1} e^k N \sum_{\ell=1}^{|\mathcal{S}|} \binom{k-1}{\ell-1} |\mathcal{S}|^{\ell} \\ & \leq \left(\binom{N}{k} |\mathcal{S}|^k \right)^{-1} e^k N |\mathcal{S}| \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} |\mathcal{S}|^{\ell} \\ & = \left(\binom{N}{k} |\mathcal{S}|^k \right)^{-1} e^k N |\mathcal{S}| (|\mathcal{S}| + 1)^{k-1} \\ & \leq ek \left(\frac{2ek}{N} \right)^{k-1} \leq \frac{2(ek)^2}{N} \left(\frac{e}{3} \right)^{k-2}. \end{aligned} \quad (11)$$

Finally, by (7), (8) and (11) we get, that for a fixed k , the probability of

$$\tilde{\Phi}_k(G) > 2\sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} \quad (12)$$

is

$$O \left(\frac{1}{N} k^2 \left(\frac{e}{3} \right)^{k-2} \right).$$

Summing it for $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$ we get that the probability that (12) holds for some k is

$$O \left(\sum_{2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)} \frac{1}{N} k^2 \left(\frac{e}{3} \right)^{k-2} \right) = O \left(\frac{1}{N} \sum_{k=0}^{\infty} k^2 \left(\frac{e}{3} \right)^k \right) = O \left(\frac{1}{N} \right).$$

Now suppose that $\log_2 N \leq \log_2 |\mathcal{S}| < N/12$. One may get in the same way, that

$$\begin{aligned} & \mathbb{P} \left(\tilde{\Phi}_k(G) > 2\sqrt{N \left(k \log N + \log \binom{|\mathcal{S}|}{k} \right)} \right) \\ & \leq \sum_{\ell} \sum_{k_1, k_2, \dots, k_{\ell}} \sum_{s_1, s_2, \dots, s_{\ell}} \sum_M \sum_D \left(N^k \binom{|\mathcal{S}|}{k} \right)^{-2}. \end{aligned} \quad (13)$$

Estimating trivially the number of terms, we get that (13) is less than

$$\left(N^k \binom{|\mathcal{S}|}{k}\right)^{-2} \sum_{\ell=1}^k \binom{k-1}{\ell-1} \binom{|\mathcal{S}|}{\ell} N^{k+1} \leq \left(N^k \binom{|\mathcal{S}|}{k}\right)^{-2} 2^{k-1} \binom{|\mathcal{S}|}{k} N^{k+1} \leq \frac{2^{k-1}}{N^{k-1} \binom{|\mathcal{S}|}{k}}.$$

Summing over $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$ we get, that the probability of (6) for some k is less than

$$N \sum_{2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)} \frac{2^k}{\binom{|\mathcal{S}|}{k} N^k} < \frac{1}{N} \sum_{k=0}^{\infty} \frac{1}{N^k} = O\left(\frac{1}{N}\right)$$

which gives the result. \square

Next, we prove the lower estimate for $\tilde{\Phi}_k(G)$ for typical generator G .

Lemma 5. *Let $m = \lfloor N/3 \rfloor$. For $1 \leq \log_2 |\mathcal{S}| \leq m^{1/4}$ we have*

$$\tilde{\Phi}_k(G) > \frac{4}{9} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)},$$

and for $m^{1/4} < \log_2 |\mathcal{S}| < N/12$ we have

$$\begin{aligned} \tilde{\Phi}_k(G) &> \frac{4}{9} \sqrt{N \left(k \log N + \log \binom{|\mathcal{S}|}{k} \right)} \\ &> \frac{4}{9} \sqrt{N \left(\log \binom{N}{k} + (1 - o(1)) k \log |\mathcal{S}| \right)}, \end{aligned}$$

with probability tending to 1 as $N \rightarrow \infty$ for every integer k with $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$.

We start with the following form of Fact 16 in [1].

Lemma 6. *Let $m = \lfloor N/3 \rfloor$. For every sufficiently large N , the followings hold.*

(i) *If $2 \leq k \leq \log m$, then*

$$\log \binom{N/3}{k} \geq 0.98 \log \binom{N}{k}.$$

(ii) *If $\log m < k \leq N/(6 \log_2 |\mathcal{S}|)$, then*

$$\log \binom{N/3}{k} \geq \frac{1 - 10^{-10}}{3} \log \binom{N}{k}.$$

Let $m = \lfloor N/3 \rfloor$ and for $1 \leq \log_2 |\mathcal{S}| \leq m^{1/4}$ consider the maximal $r = r_k(m, \mathcal{S}) \in \mathbb{N}$ such that

$$\mathbb{P} \left(S(m, 1/2) \geq \frac{1}{2}(m + r) \right) \geq \frac{k^2 \log N}{\binom{m+1}{k-1} |\mathcal{S}|^k}$$

holds, and for $m^{1/4} < \log_2 |\mathcal{S}| \leq N/12$ consider the maximal $r = r_k(m, \mathcal{S}) \in \mathbb{N}$ such that

$$\mathbb{P} \left(S(m, 1/2) \geq \frac{1}{2}(m + r) \right) \geq \frac{k^2 \log N}{(m+1)^{k-1} \binom{|\mathcal{S}|}{k}}$$

holds.

We give a lower estimate to $r_k(m, \mathcal{S})$ for large and small \mathcal{S} separately.

Lemma 7. *For every sufficiently large N and for $1 \leq \log_2 |\mathcal{S}| \leq m^{1/4}$ the followings hold.*

(i) *For $2 \leq k \leq \log m$ we have*

$$r_k(m, \mathcal{S}) \geq 0.99 \sqrt{2m \left(\log \binom{m+1}{k-1} + k \log |\mathcal{S}| \right)}.$$

(ii) *For $\log m < k \leq N/(6 \log_2 |\mathcal{S}|)$ we have*

$$r_k(m, \mathcal{S}) \geq (1 - 10^{-10}) \sqrt{\frac{1}{\log 2} m \left(\log \binom{m+1}{k-1} + k \log |\mathcal{S}| \right)}.$$

(iii) *For $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$ we have*

$$r_k(m, \mathcal{S}) \geq \frac{4}{9} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)}.$$

Lemma 7. First we remark that for all $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$, we have

$$k^2 \log N \leq \binom{m+1}{k-1}^{o(1)},$$

(see e.g. [1]).

First assume, that $k \leq \log m$. Let

$$r = \left\lceil 0.99 \sqrt{2m \left(\log \binom{m+1}{k-1} + k \log |\mathcal{S}| \right)} \right\rceil$$

and

$$c = \frac{r+1}{2\sqrt{m}} = (1 + o(1)) 0.99 \sqrt{\frac{1}{4} \left(\log \binom{m+1}{k-1} + k \log |\mathcal{S}| \right)}.$$

Since now $c = o(m^{1/6})$, by (i) of Lemma 1 we have

$$\begin{aligned} \mathbb{P}\left(S(m, 1/2) \geq \frac{1}{2}(m+r)\right) &\geq \mathbb{P}\left(S(m, 1/2) \geq \left\lfloor \frac{m}{2} \right\rfloor + c\sqrt{m}\right) \\ &= \frac{e^{-2c^2}}{2c\sqrt{2\pi}}(1+o(1)) \geq \frac{1}{4} \left(\binom{m+1}{k-1} |\mathcal{S}|^k \right)^{-0.99} \geq \frac{k^2 \log N}{\binom{m+1}{k-1} |\mathcal{S}|^k} \end{aligned}$$

which proves (i).

To prove (ii) assume, that $\log m < k \leq N/(6 \log_2 |\mathcal{S}|)$. Let

$$r = \left\lceil (1 - 10^{-10}) \sqrt{\frac{1}{\log 2} m \left(\log \binom{m+1}{k-1} + k \log |\mathcal{S}| \right)} \right\rceil$$

and

$$c = \left\lceil \frac{r+1}{2} \right\rceil = (1+o(1)) \frac{1-10^{-10}}{2\sqrt{\log 2}} \sqrt{m \left(\log \binom{m+1}{k-1} + k \log |\mathcal{S}| \right)}.$$

Since now $0 < c < m/2$, by Lemma 2 we have

$$\begin{aligned} \mathbb{P}\left(S(m, 1/2) \geq \frac{1}{2}(m+r)\right) &\geq \mathbb{P}\left(S(m, 1/2) \geq \left\lfloor \frac{m}{2} \right\rfloor + c\right) \\ &\geq (1+o(1)) 2^{-4(c+1/2)^2/m} \sqrt{\frac{2}{\pi m}} \geq (1+o(1)) 2^{-r^2/m} 2^{-(6r+9)/m} \sqrt{\frac{2}{\pi m}} \\ &\geq \left(\binom{m+1}{k-1} |\mathcal{S}|^k \right)^{-1+10^{-10}} 2^{-(6r+9)/m} \sqrt{\frac{2}{\pi m}} \\ &\geq (1+o(1)) \left(\binom{m+1}{k-1} |\mathcal{S}|^k \right)^{-1+10^{-10}} \geq \frac{k^2 \log N}{\binom{m+1}{k-1} |\mathcal{S}|^k}. \end{aligned}$$

Finally, (iii) follows from (i), (ii) and Lemma 6 in the same way as in [1]. Namely, if $2 \leq k \leq \log m$, then

$$\binom{m+1}{k-1} \geq \binom{N/3}{k-1} \geq \binom{N/3}{k}^{1/2},$$

thus

$$\begin{aligned} r_k(m, \mathcal{S}) &\geq 0.99 \sqrt{2 \left\lfloor \frac{N}{3} \right\rfloor \left(\log \binom{m+1}{k-1} + k \log |\mathcal{S}| \right)} \\ &\geq (1+o(1)) \frac{0.99}{\sqrt{3}} \sqrt{N \left(\log \binom{N/3}{k} + k \log |\mathcal{S}| \right)} \\ &\geq \frac{4}{9} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)}. \end{aligned}$$

On the other hand, if $\log m < k \leq N/(6 \log_2 |\mathcal{S}|)$, then

$$\binom{m+1}{k-1} \geq \left(\frac{N}{3}\right)^{1-o(1)},$$

thus

$$\begin{aligned} r_k(m, \mathcal{S}) &\geq \frac{1-10^{-10}}{\sqrt{\log 2}} \sqrt{\left\lfloor \frac{N}{3} \right\rfloor \left(\log \binom{m+1}{k-1} + k \log |\mathcal{S}| \right)} \\ &\geq (1+o(1)) \frac{1-10^{-10}}{\sqrt{3 \log 2}} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)} \\ &\geq \frac{4}{9} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)}. \end{aligned}$$

□

The lower estimate to $r_k(m, \mathcal{S})$ for small \mathcal{S} can be prove similarly.

Lemma 8. *For every sufficiently large N and for $m^{1/4} < \log_2 |\mathcal{S}| < N/12$*

$$r_k(m, \mathcal{S}) \geq \frac{4}{9} \sqrt{N \left(\log \binom{N}{k} + k \log |\mathcal{S}| \right)}.$$

holds for $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$.

We also need the following lemma ([1, Lemma 19]).

Lemma 9. *Let A_1, A_2, \dots, A_M be events in a probability space, each with probability at least p . Let $\varepsilon \geq 0$ be given, and suppose that*

$$\mathbb{P}(A_i \cap A_j) \leq p^2(1 + \varepsilon)$$

for all $i \neq j$. Then

$$\mathbb{P} \left(\bigcup_{i=1}^M A_i \right) \geq 1 - \varepsilon - \frac{2}{Mp}.$$

Now we are in the state to prove Lemma 5.

Lemma 5. First we remark, that it is enough to show that

$$\tilde{\Phi}_k(G) \leq r_k(m, \mathcal{S}) \tag{14}$$

holds with probability at most $O(1/k^2 \log N)$.

Indeed, summing over all $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$ we get that (14) holds for *some* k with $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$ with probability $O(1/\log N) = o(1)$. Whence (14) does not hold for all $2 \leq k \leq N/(6 \log_2 |\mathcal{S}|)$ with probability $1 - o(1)$, which proves the lemma.

We prove the lemma for small \mathcal{S} , for large \mathcal{S} one can obtain the result in the same way referring to Lemma 8 instead of Lemma 7. So assume, that $1 \leq \log_2 |\mathcal{S}| \leq m^{1/4}$.

For $s_1 \in \mathcal{S}$ let

$$v(s_1) = (e_1(s_1), e_2(s_1), \dots, e_m(s_1))$$

and for $2 \leq \ell \leq k$, for $s_2, \dots, s_\ell \in \mathcal{S}$, for $(k_2, k_3, \dots, k_\ell)$ with $k_2 + k_3 + \dots + k_\ell = k - 1$ and for $D = (0, d_1^2, d_2^2, \dots, d_{k_2}^2, \dots, d_1^\ell, d_2^\ell, \dots, d_{k_\ell}^\ell)$ with $m \leq d_1^j < \dots < d_{k_\ell}^j \leq 2m$ for $j = 2, 3, \dots, \ell$ let

$$v_\ell(s_2, s_3, \dots, s_\ell, D) = \left(\prod_{t=2}^{\ell} e_{1+d_1^t}(s_t) \cdots e_{1+d_{k_t}^t}(s_t), \dots, \prod_{t=2}^{\ell} e_{m+d_1^t}(s_t) \cdots e_{m+d_{k_t}^t}(s_t) \right)$$

Let $A_\ell(s_1, s_2, \dots, s_\ell, D)$ be the event

$$|\langle v(s_1), v_\ell(s_2, s_3, \dots, s_\ell, D) \rangle| \geq r_k(m, \mathcal{S}).$$

Since $\langle v(s_1), v_\ell(s_2, s_3, \dots, s_\ell, D) \rangle$ has the same distribution as $S(m, 1/2)$, we have

$$p = \mathbb{P}(A_\ell(s_1, s_2, \dots, s_\ell, D)) = 2 \cdot \mathbb{P}\left(S(m, 1/2) \geq \frac{1}{2}(m + r_k(m, \mathcal{S}))\right).$$

One can obtain in the same way as [1, Claim 18], that the events A_ℓ are pairwise independent.

Lemma 10. For $\{s_1, s_2, \dots, s_\ell\} \neq \{s'_1, s'_2, \dots, s'_{\ell'}\}$ or $D \neq D'$ we have

$$\mathbb{P}(A_\ell(s_1, s_2, \dots, s_\ell, D) \cap A_{\ell'}(s'_1, s'_2, \dots, s'_{\ell'}, D')) = p^2.$$

Let $\mathcal{D}_k(\mathcal{S})$ be the number of possible ℓ , s_1, s_2, \dots, s_ℓ and D , then by Lemmas 9 and 10 we get that

$$\begin{aligned} & \mathbb{P}\left(\tilde{\Phi}_k(G) \geq r_k(m, \mathcal{S})\right) \\ & \geq \mathbb{P}\left(\bigcup_{\ell=2}^k \bigcup_{\substack{k_2, \dots, k_\ell \geq 1 \\ k_2, \dots, k_\ell = k-1}} \bigcup_{s_1, s_2, \dots, s_\ell} \bigcup_D A_\ell(s_1, s_2, \dots, s_\ell, D)\right) \geq 1 - \frac{2}{p \cdot \mathcal{D}_k(\mathcal{S})}. \end{aligned} \quad (15)$$

Finally, we give a lower bound to (15) for $|\mathcal{S}| < m$ and $|\mathcal{S}| \geq m$ separately. If $|\mathcal{S}| < m$, then

$$\begin{aligned}
p \cdot \mathcal{D}_k(\mathcal{S}) &= p \sum_{\ell=2}^k \sum_{\substack{k_2, \dots, k_\ell \geq 1 \\ k_2, \dots, k_\ell = k-1}} \sum_{s_1, s_2, \dots, s_\ell} \sum_D 1 \\
&= p \sum_{\ell=2}^k \sum_{\substack{k_2, \dots, k_\ell \geq 1 \\ k_2, \dots, k_\ell = k-1}} \sum_{s_1, s_2, \dots, s_\ell} \binom{m+1}{k_2} \binom{m+1}{k_3} \cdots \binom{m+1}{k_\ell} \\
&\geq p \sum_{\ell=2}^k \binom{k-2}{\ell-2} \binom{|\mathcal{S}|}{\ell} m^{\ell-2} \binom{m+1}{k-1} \\
&\geq p \binom{|\mathcal{S}|}{2} \binom{m+1}{k-1} \sum_{\ell=2}^k \binom{k-2}{\ell-2} |\mathcal{S}|^{\ell-2} \\
&\geq \frac{1}{4} p \binom{m+1}{k-1} (|\mathcal{S}|)^k \geq \frac{1}{2} k^2 \log N
\end{aligned}$$

Similarly, for $|\mathcal{S}| \geq m$ we have

$$\begin{aligned}
p \cdot \mathcal{D}_k(\mathcal{S}) &= p \sum_{\ell=2}^k \sum_{\substack{k_2, \dots, k_\ell \geq 1 \\ k_2, \dots, k_\ell = k-1}} \sum_{s_1, s_2, \dots, s_\ell} \sum_D 1 \\
&\geq p \sum_{s_1, s_2, \dots, s_k} \binom{m+1}{1} \cdots \binom{m+1}{1} = p \binom{|\mathcal{S}|}{k} (m+1)^{k-1} \geq \frac{1}{2} k^2 \log N
\end{aligned}$$

which proves the result. \square

Acknowledgements

The author is partially supported by the Austrian Science Fund FWF Project F5511-N26 which is part of the Special Research Program "Quasi-Monte Carlo Methods: Theory and Applications" and by Hungarian National Foundation for Scientific Research, Grant No. K100291.

References

- [1] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, Measures of pseudo-randomness for finite sequences: typical values, Proc. Lond. Math. Soc. (3) 95 (2007), no. 3, 778–812.

- [2] N. Alon and J. H. Spencer, The probabilistic method, second ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience [John Wiley & Sons], New York, 2000, With an appendix on the life and work of Paul Erdős
- [3] J. Cassaigne, C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences VII: The measures of pseudorandomness, *Acta Arith.* 103 (2002), no. 2, 97–118.
- [4] B. Bollobás, Random graphs, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985.
- [5] K. Gyarmati, Measures of pseudorandomness, P. Charpin, A. Pott, A. Winterhof (eds.), Radon Series in Computational and Applied Mathematics, de Gruyter 2013, 43-64.
- [6] K. Gyarmati, C. Mauduit, A. Sárközy, The cross-correlation measure for families of binary sequences, *Applications of Algebra and Number Theory* (Lectures on the occasion of Harald Niederreiter's 70th Birthday)
- [7] C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences I: Measures of pseudorandomness, the Legendre symbol, *Acta Arith.* 82 (1997) 365–377.
- [8] A. J. Menezes, P. C. Oorschot, S. A. Vanstone, Handbook of applied cryptography, CRC Press Series on Discrete Mathematics and its Applications. CRC Press, Boca Raton, FL, 1997
- [9] K.-U. Schmidt, The correlation measures of finite sequences: limiting distributions and minimum values, *Trans. Amer. Math. Soc.*, to appear.